Explainable Recommendation with Review Generation

Group 5-2:
Wout Kooijman, Kai Liang, Weitao Luo, Erik Stammes, Ozzy Ülger, Vicky Foing

A. Introduction

- Explainability is key to successful recommender systems
- Online user reviews can be used a source of information
  a. Extract sentiments and aspects
  b. Generating synthetic reviews
- Research goals: Rating prediction & Review generation
  a. Reproduce baseline results
  b. Extend baseline models to obtain better results

A.1 Explainable Recommendations

- Types of explanations
  - User/item, content, **textual**, visual, and social explanations
- Models
  - **MF**, topic, graph, **deep learning**, knowledge, mining, post-hoc
- Evaluation
  - RMSE, offline/online evaluation, user study, qualitative/quantitative
- Applications
  - **E-commerce**, POI, social, multi-media recommendations
- Future directions
  - Knowledge graphs, DL, NLP, user behavior analysis, logic, cognitive foundations

B. Related Work

- DeepCoNN
- DER
- CTRL
- MTER*
- MRG*
- NARRE*

B.1 DeepCoNN[2]

- Computes rating that user would assign to item
- Model: CNN consisting of two parallel neural networks coupled in the last layers
    a. Learning user behaviors from the user's reviews
    b. Learning item properties from the item's reviews
- Performance (2017):
    a. Better results for rating prediction than MF, topic, and DL-based models
    b. Better results than MF when there is data sparsity
- Limitations:
    a. Performs best when review for items is present in dataset. Not like real world[1]

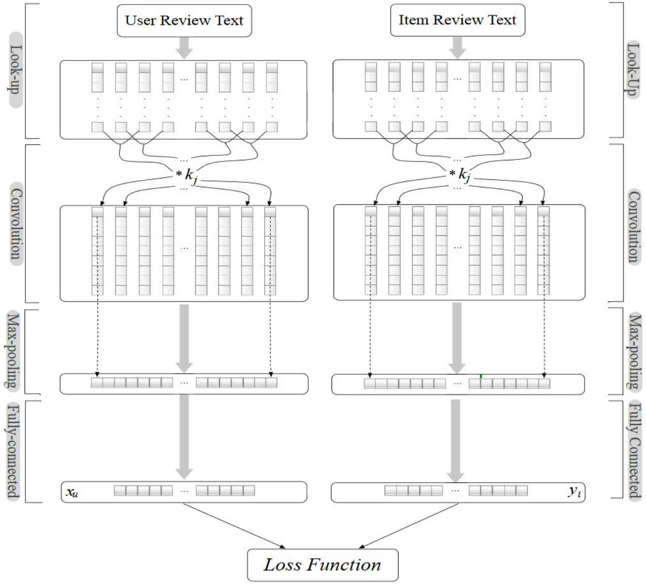
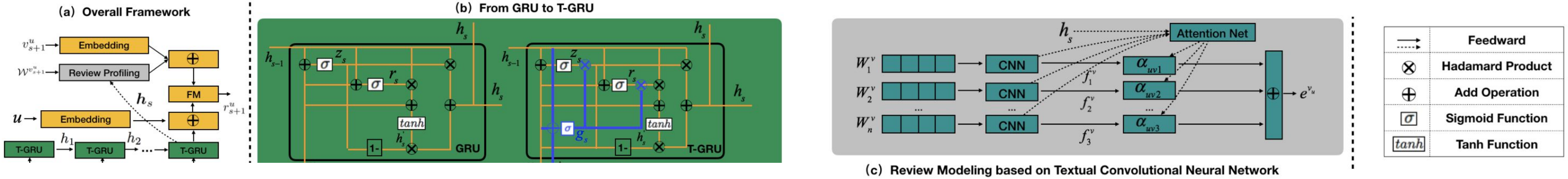


Figure 1: The architecture of the proposed model

B.2 Dynamic Explainable Recommendations (DER)[4]

- Computes rating that user would assign to item
- Provides explanation by highlighting important sentences in the item's reviews
- Model: GRU + CNN + attention mechanism
  - A time-aware GRU models user's dynamic preferences
  - A sentence-level CNN extracts item properties from reviews
- Performance (2019):
  - Better rating prediction than NARRE (which outperforms DeepCoNN)



(c) Review Modeling based on Textual Convolutional Neural Network

B.3 Conditional Transformer Language Model for Controllable Generation (CTRL)

- A conditional transformer for controllable language generation
- Trained to condition on control codes that specify domain, subdomain, entities and more
- Can generate wikipedia articles, news articles, translation, **reviews** and more

`Links https://www.cnn.com/2014/09/20/us-president-meets-british-pm`
JUST WATCHED\n\nObama meets with British PM\n\nMUST WATCH\n\nStory highlights\n\nPresident Barack Obama met with Britain's Prime Minister David Cameron

`Links https://www.cnn.com/2018/09/20/us-president-meets-british-pm`
JUST WATCHED\n\nTrump and May meet for first time\n\nMUST WATCH\n\nWashington (CNN) President Donald Trump, who has been criticized by some in the UK over his decision to leave the European Union, met with British Prime Minister Theresa May, a White House official said on Thursday.

B.4 Multimodal Review Generation (MRG)[3]

- Model: A neural approach with two components:
    a. Rating prediction
    b. Review text generation
- Can use images to inform review text generation
- Performance (2019):
    a. Better rating prediction than MF models
    b. Better review text generation than LSTM models

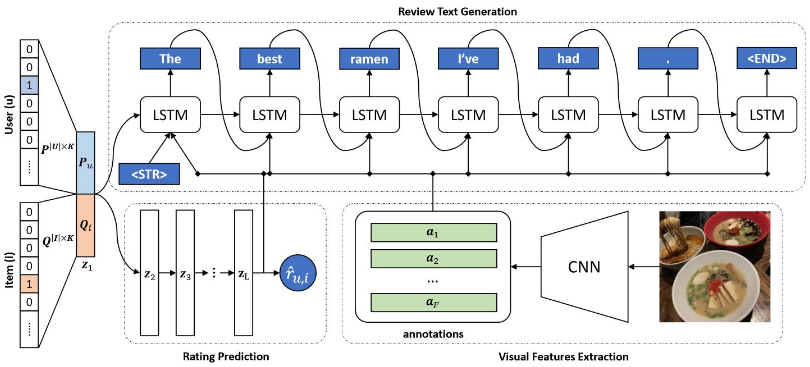


Figure 2: Overall Architecture of Multimodal Review Generation (MRG) model

C. Experiments

- Models
- Datasets
- Evaluation

C.1 Reproduction Results

All models are trained on the small 5-core Amazon topic datasets[6] and are evaluated using mean squared error (MSE)

	Digital Music
DeepCoNN (paper)	0.749* - Trained on full dataset
DeepCoNN (ours)	0.783

	Musical Instruments	Automotive	Toy
DER (paper)	0.9678	0.8981	0.9535
DER (ours)	0.9044	0.9022	0.9317

C.2.1 CTRL[5]

- More fine-grained control codes
  - Use Amazon dataset to include product properties in control codes
- Training code not (yet) available
- Combination of control codes is already possible

```
Reviews Rating: 1.0    Apple iPhone X 64GB Space Gray
```

```
Links https://www.amazon.de/HPE-NVIDIA-Tesla-Dual-Module
```

C.2.2 CTRL[5]

- Reviews Rating: 1.0 Apple iPhone X 64GB Space Gray (Wireless Phone Accessory)I bought this case for my new iphone and it was a waste of money. The plastic is cheap and the rubber part does not fit well at all. It also makes your screen look blurry.

- Links https://www.amazon.com/Marvels-Spider-Man-PlayStation-4 Reviews Rating: 4.0 If you're looking at this, then you've probably already played Spider-Man 2 or 3. So I'll skip over those two games, because they're pretty much identical except for their different endings. In fact, there isn't really anything new about them besides the fact that they've added more levels and characters. But that's okay. Because what makes these games unique is how well they integrate into each other.

D. Methods/Ambition

Build upon the MRG model with improvements in

- Rating prediction: classification instead of regression
- Final output layer: improve the LSTM's output

Why MRG?

- Unifies two tasks in a simple, extensible architecture
- Most recent model with available code
- It outperforms both content-based recommendation systems and review text generation baselines

D.1 Alter Rating Prediction's Architecture

To the best of our understanding, most baselines in the field treat rating prediction as a regression task by using MLP

What if we treat prediction as a classification task instead

Motivations:

- The domain of rating is not the whole real number field in any online selling system anyway
  - E.g. you cannot grade an item by 3.1415926, -1 or 10000
  - Thus, the nature of the problem is more a classification task (e.g. integers from 1 to 5 or 10)
- Treating rating prediction as a classification task allows more NN architecture options (e.g. CNN)

Hypothesis: By adopting the nature of the problem, and using more complex architectures, the rating prediction will be more accurate

D.2.1 Improve LSTM's Output

Reviews in the training set often include relatively simple words used by humans, such as

"The food here is ok but not worth the price."

A recently proposed "neural editor" model enables semantic improvement such that the above becomes

"The food is mediocre and not worth the ridiculous price."

Our idea is to use a neural editor model to

1. The output layer of the MRG LSTM in order to generate more qualitative sentences
2. Invoke the neural editor model after each full stop.

Hypothesis: The prototype editing model will improve the semantic quality of the generated output of the MRG model

D.2.2 Improve LSTM's Output - Idea 1

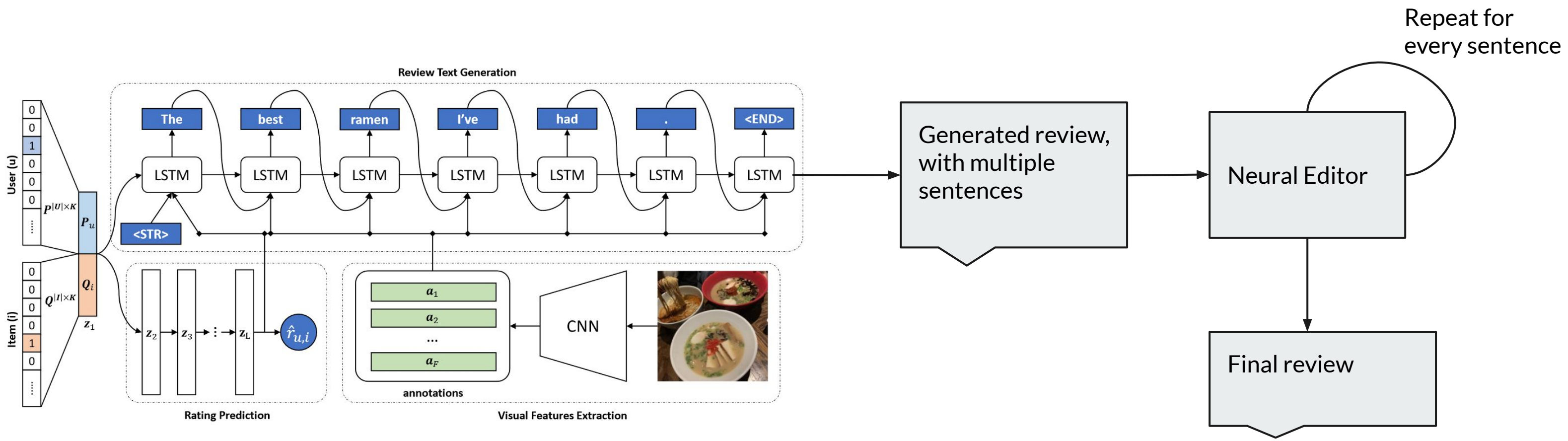


Figure 2: Overall Architecture of Multimodal Review Generation (MRG) model

D.2.3 Improve LSTM's Output - Idea 2

After each full stop, invoke neural editor model and restart training

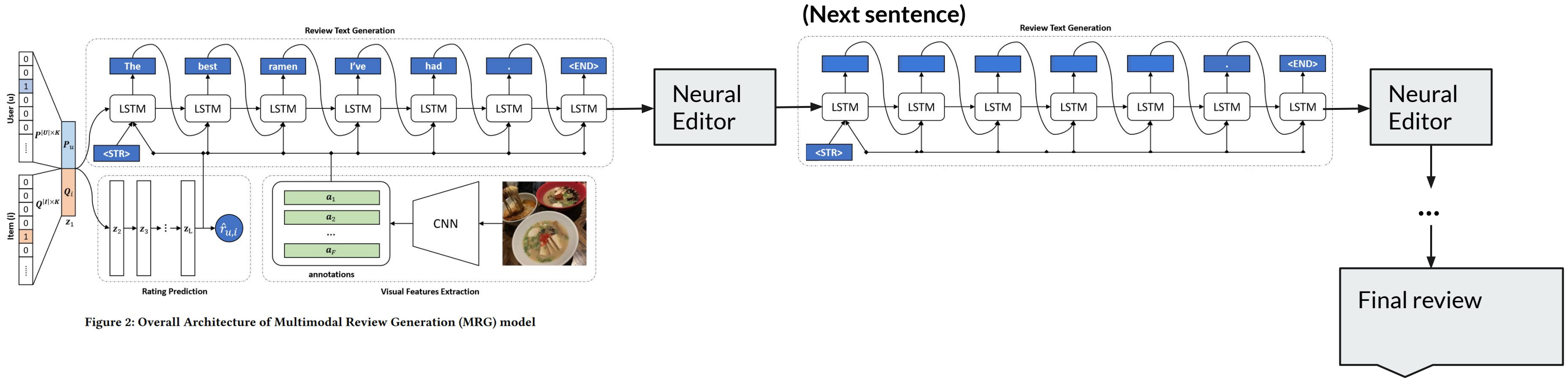


Figure 2: Overall Architecture of Multimodal Review Generation (MRG) model

E. Final Evaluation

Measuring semantic quality:

- BLEU (**precision**): How much does the generated review overlap with human review
- ROUGE (**recall**): How much does the human review overlap with generated review

Measuring rating prediction:

- RMSE: **regression**
- Cross-entropy loss: **classification**

If time permits, conduct ablation analysis to investigate the contribution of architecture components

References

1) Kanjirathinkal, Rose Catherine. *Explainable Recommendations*. Diss. Stanford University, 2018.
2) Zheng, Lei, Vahid Noroozi, and Philip S. Yu. "Joint deep modeling of users and items using reviews for recommendation." *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 2017.
3) Truong, Quoc-Tuan, and Hady Lauw. "Multimodal Review Generation for Recommender Systems." *The World Wide Web Conference*. ACM, 2019.
4) Chen, Xu, Yongfeng Zhang, and Zheng Qin. "Dynamic Explainable Recommendation based on Neural Attentive Models." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019.
5) Keskar, Nitish Shirish, et al. "CTRL: A Conditional Transformer Language Model for Controllable Generation." *arXiv preprint arXiv:1909.05858* (2019).
6) R. He, J. McAuley "Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering." *WWW*, 2016